

WE CLAIM:

1. A security service for a shared network server comprising:

5 providing a network and a shared network server coupled to the network, the shared network server having a fixed quantity of resources for responding to network requests;

providing a constellation of front-end servers within the network;

10 using the front-end servers to receive requests destined for the shared network server; and

forwarding the received requests from the front-end servers to the shared network server at a governed rate.

2. The service of claim 1 wherein the governed rate is selected to present requests at a rate that will prevent overwhelming the fixed quantity of resources within the shared network server.

3. The service of claim 1 further comprising coupling a management server to each of the front-end servers;

5 communicating metrics between the front-end servers and the management server; and

using the metrics to detect a denial of service attack targeted at the shared network server.

4. The service of claim 3 further comprising:

using the metrics to determine configuration parameters for the front-end servers; and

5 communicating the configuration parameters from the management server to the front-end servers.

5. The service of claim 1 further comprising dynamically altering the number of front-end servers in the constellation.

6. The service of claim 1 further comprising detecting a denial of service attack targeted at the shared network server; and

5 preventing the act of forwarding the received requests in response to detecting the DoS attack.

7. A system for handling denial of service attacks on behalf of a shared network resource, the system comprising:

5 a request processing component deployed within a network, the request processing component having an interface configured to receive requests on behalf of the shared network server;

10 a rate control component coupled to the request processing component, the rate control component comprising program and data structures operable to selectively forward received requests to the shared network server at a rate selected to prevent the shared network server from crashing or becoming undesirably busy.

8. A system of claim 7 further comprising:

a DoS attack detection component coupled to the request processing component and the rate control component and operable to monitor request metrics from 5 the request processing component and provide configuration information to the rate control component.

9. The system of claim 8 wherein the rate control component comprises mechanisms for preferentially forwarding requests not related to the DoS attack in favor of request related to the DoS attack to the shared network resource.

10. The system of claim 7 further comprising:

a plurality of front-end servers deployed throughout a network, wherein the front-end servers are configured to implement the request processing component and the
5 rate control component;

a management server coupled to each of the front-end servers, the management server including mechanisms to send configuration information to the front-end servers. and receive request processing metrics from the request
10 processing component.

11. The system of claim 7 wherein the request processing component is configured to handle a greater volume of requests than the shared network resource.

12. The system of claim 7 further comprising mechanisms within the front-end servers operable to detect a denial of service attack; and

13. The system of claim 10 further comprising a back-end server coupled to receive the forwarded requests from the front-end servers; and

5 a rate governor within the back-end server for selectively forwarding received requests to the shared network resource at a rate selected to prevent the shared network resource from crashing becoming undesirably busy.

14. A method for mitigating a denial of service attack comprising the acts of:

providing a shared network resource coupled to a public network and receiving requests from the public
5 network;

providing a plurality of front-end servers, each having a unique network address and coupled to the shared network resource;

10 assigning a plurality of front-end servers to the shared network resource, wherein the aggregate request

processing capacity of the assigned front-end servers is greater than the request handling capacity of the shared network resource;

15 causing requests for the shared network resource to be redirected through one of the front-end servers; and

forwarding the requests from the front-end server to the shared network resource at a rate selected to inhibit a likelihood of a crash or an undesirable level of business.

15. The method of claim 14 further comprising:

in event of a denial of service attack comprising a plurality of malicious requests involving the shared network resource, causing at least some of the malicious requests to be delayed in the front-end servers before reaching the shared network resource.

16. The method of claim 14 further comprising:

in event of a denial of service attack comprising a plurality of malicious requests involving the shared network resource, causing at least some of the malicious requests to be ignored by the front-end servers.

17. The method of claim 15 further comprising:

in event of a denial of service attack comprising a plurality of malicious requests involving the shared network resource, causing at least some of the malicious requests to be ignored in the front-end servers before reaching the shared network resource.

18. The method of claim 14 further comprising acts of:

detecting a condition in which the number of requests is greater than the request capacity of the shared network resource; and

generating a response to the requests from the front-end servers instead of forwarding the requests to the shared network resource.

19. The method of claim 17 wherein the act of detecting comprises distinguishing requests associated with a DoS attack from legitimate requests and the step of generating a response comprises generating a response only to requests associated with the DoS attack while forwarding legitimate requests to the shared network resource.

20. The method of claim 14 further comprising:
sending request processing metrics from each of the front-end servers to a centralized management server; and
using the centralized management server to analyze
5 the request processing metrics to detect a denial of service attack.

21. The method of claim 19 further comprising:
sending configuration information from the centralized management server to some or all of the front-end servers in response to detecting a DoS attack,
5 the configuration information including an identification of address domains associated with the DoS attack; and
using the configuration information in the front-end server to selectively drop requests from the address domain identified in the configuration information.